

Université de Montpellier

Département Hérault

Mini-Projet de Big Data

Analyse décisionnelle et conception d'un entrepôt de données



Étudiants

Grosu Andy
Guncavdi Ay
Sireix Alexandre

Encadrants

Ens. Arigon Anne Muriel
Ens. Ulliana Federico

Année universitaire : 2025–2026

Montpellier, le 02/11/2025

Table des matières

1	Question 1 - Analyse des besoins métier	2
1.1	Les chiffres d'affaires en 2024	3
1.2	Infos importantes	3
1.3	Besoins métiers principaux	3
2	Question 2 - Étude des actions principales	3
2.1	Action 1 : Étude des ventes	3
2.2	Action 2 : Étude du stock	4
2.3	Action 3 : Étude de la fiabilité (retours au garage)	4
3	Question 3 – Objectifs, actions et traitements analytiques	4
4	Question 4 - Ordre d'importance	5
5	Question 5 - Actions les plus importantes	5
6	Question 7 - Conception d'un datamart	6
6.1	Question 9 - Liste des mesures	6
6.2	Question 10 - Raisonnement pour répondre aux besoins métier	6
6.3	Question 11- Pertinence des modèles	6
6.4	Question 12 - Vérification	7
7	Question 13 - Snapshot	8
7.1	Question 16 - Liste des mesures	8
7.2	Question 17 - Raisonnement pour répondre aux besoins métier	8
7.3	Question 18 - Pertinence des modèles	8
7.4	Question 19 - Estimation de la taille du datamart	10
8	Question 14 - Updated record	10
8.1	Question 16 - Liste des mesures	10
8.2	Question 17 - Raisonnement pour répondre aux besoins métier	10
8.3	Question 18 - Pertinence des modèles	11
8.4	Question 19 - Estimation de la taille du datamart	11

1 Question 1 - Analyse des besoins métier

Questions directrices

- Identifier les objectifs de l'entreprise (ou institution) choisie.
- Déterminer sa position sur le marché.
- Présenter ses produits et services, ainsi que ses sources de revenu.
- Identifier les informations utiles pour la prise de décision interne.
- Indiquer les actions/opérations à tracer pour obtenir ces informations.
- Pour chaque action/opération, proposer au moins trois traitements possibles (requêtes analytiques).
- Montrer comment chaque requête contribue à un besoin métier.
- Classer les actions par importance ou rentabilité potentielle.
- Identifier les deux actions/opérations les plus importantes à analyser.

Présentation de l'entreprise

Le groupe **Volkswagen** a plusieurs objectifs courants, d'une part d'augmenter ses recettes et d'autre part de rembourser ses dettes. Il était le **deuxième plus grand producteur de véhicules au monde** en termes de vente, avec une production de **9 030 000 articles** et un **chiffre d'affaires de 324 656 milliards en 2024**.

Le groupe propose plusieurs produits et services, notamment : *automobiles, véhicules commerciaux, moteurs à combustion interne, motos, camions, camionnettes, turbomachines, services bancaires, financement, gestion de flotte, assurance et location*.

Cela est possible car le groupe est en possession des marques suivantes :

- Volkswagen (Allemagne)
- Audi (Allemagne)
- Seat (Espagne)
- Skoda (République tchèque)
- Bugatti (France)
- Lamborghini (Italie)
- Bentley (Angleterre)
- Volkswagen Nutzfahrzeuge (Allemagne)
- MAN (Allemagne)
- Ducati (Italie)
- Porsche (Allemagne)
- Scania et Saab Automobile (Suède)
- Scout (États-Unis)

Il intervient dans **153 pays du monde**, avec une centaine d'usines réparties à travers **27 pays**. Il est dans une dette considérable de **155,6 milliards d'euros**, et doit donc trouver des moyens de faire des économies sans provoquer une chute considérable de ses revenus.

Une stratégie est de s'informer sur les risques du marché chaque année. *E.g. Y'a-t-il des pénuries dans certaines régions du monde, notamment là où le groupe dispose des usines ? Si c'est le cas, comment s'assurer que la production des articles ne soit pas impactée ?*

1.1 Les chiffres d'affaires en 2024

Ventes par service

Sector	Share
Voitures particulières et véhicules utilitaires légers	68,9%
Services financiers	16,7%
Véhicules utilitaires	13,1%
Ingénierie énergétique	1,2%

Ventes par région

Region	Share
Europe / Autres marchés	40,6%
Amérique du Nord	20,9%
Allemagne	19,1%
Asie-Pacifique	13,6%
Amérique du Sud	5,8%

1.2 Infos importantes

Les informations les plus importantes pour le groupe sont d'abord leurs ventes par service et par région. Cela est essentiel pour que le groupe puisse discerner où faire ses dépenses. Par exemple, il pourrait décider de fermer des usines dans des régions les moins profitables. Ensuite, la performance de leurs usines, qui est aussi vitale à la prise de décision de leurs actionnaires. Enfin, un dernier point dominant qui est les tendances de consommations de ses clients. E.g. Historiquement, le modèle le plus rentable du groupe est le Golf, qui a vendu plus de 35 millions d'unités en 2019. Ça peut être sensé de se concentrer sur la production de ce dernier.

1.3 Besoins métiers principaux

- Augmenter les recettes.
- Optimiser la gestion du stock de véhicules.
- Améliorer la fiabilité des voitures vendues.

2 Question 2 - Étude des actions principales

2.1 Action 1 : Étude des ventes

Objectif : Augmenter le bénéfice global.

Traitements proposés :

- Nombre de ventes en fonction de la marque, du pays, de la ville, du profil client ou du caractère *seconde main*.
- Coût des ventes en fonction des marques.
- Ventes par gamme, modèle, ou concessionnaire.
- Analyse des ventes par promotions et par période.

2.2 Action 2 : Étude du stock

Objectif : Améliorer la gestion et la disponibilité du stock.

Traitements proposés :

- Nombre de véhicules restants en stock par ville, région, pays, marque et catégorie.
- Quantité de voitures produites par mois, par modèle et par gamme.
- Analyse des coûts de production par marque et par période.

2.3 Action 3 : Étude de la fiabilité (retours au garage)

Objectif : Optimiser la satisfaction client et la gestion de l'entretien.

Traitements proposés :

- Nombre de véhicules retournés au garage selon la marque et la pièce défectueuse.
- Quantité de réparations par type de pièce et par modèle.
- Coût de réparation par marque et par année de production.

3 Question 3 – Objectifs, actions et traitements analytiques

Objectif 1 : Augmenter le bénéfice

Action 1 : Étudier les ventes

- **Requête 1 : Analyse du nombre de ventes selon différents critères**
Nombre de véhicules vendus par *marque, pays, ville, profil client* et *état (neuf ou seconde main)*.
Utilité métier : identifier les zones géographiques et les segments de clientèle les plus rentables, afin de concentrer les efforts marketing sur ces cibles.
- **Requête 2 : Étude du coût des ventes par segment**
Coût moyen et total des ventes par *marque, gamme* et *modèle*.
Utilité métier : repérer les produits les plus coûteux à vendre et ajuster les marges pour maximiser le bénéfice global.
- **Requête 3 : Analyse des ventes selon les promotions et les concessionnaires**
Volume de ventes par *promotion, gamme* et *concessionnaire*.
Utilité métier : mesurer l'impact réel des campagnes promotionnelles et des performances des concessionnaires sur le chiffre d'affaires.

Objectif 2 : Améliorer la gestion du stock

Action 2 : Étudier le stock

- **Requête 1 : Suivi des véhicules disponibles**
Nombre de voitures restantes en stock par *ville, région, pays, marque* et *catégorie*.
Utilité métier : éviter la surproduction ou les ruptures de stock et optimiser la répartition des véhicules dans les différentes zones géographiques.

- **Requête 2 : Suivi de la production mensuelle**
Quantité de véhicules produits par *mois*, *modèle* et *gamme*.
Utilité métier : ajuster la production selon la demande réelle observée afin de réduire les coûts de stockage et les invendus.
- **Requête 3 : Analyse du ratio production/vente**
Comparaison entre le nombre de véhicules produits et le nombre de véhicules vendus par *gamme*.
Utilité métier : détecter les écarts entre production et ventes et adapter les stratégies industrielles.

Objectif 3 : Suivre la fiabilité des véhicules et optimiser la satisfaction client

Action 3 : Étudier les retours au garage

- **Requête 1 : Taux de retour par marque et type de pièce**
Nombre de véhicules revenus en réparation par *marque* et *pièce remplacée*.
Utilité métier : identifier les marques ou composants les plus défaillants et améliorer la qualité de production.
- **Requête 2 : Volume et coût des réparations**
Quantité d'interventions et coût total des réparations par *pièce* et *modèle de voiture*.
Utilité métier : obtenir une vision précise du coût après-vente et optimiser les contrats de garantie ou les accords fournisseurs.
- **Requête 3 : Analyse du coût de réparation selon l'année de production**
Coût moyen de réparation par *année de production* et *type de véhicule*.
Utilité métier : détecter les générations de modèles moins fiables et planifier des améliorations techniques sur les nouvelles séries.

4 Question 4 - Ordre d'importance

L'ordre d'importance des actions identifiées est présenté ci-dessous :

1. Étude des ventes
2. Étude de la fiabilité
3. Optimisation des stocks

5 Question 5 - Actions les plus importantes

Parmi les actions précédemment citées, les plus importantes à analyser sont :

1. Étude des ventes
2. Étude de la fiabilité

6 Question 7 - Conception d'un datamart

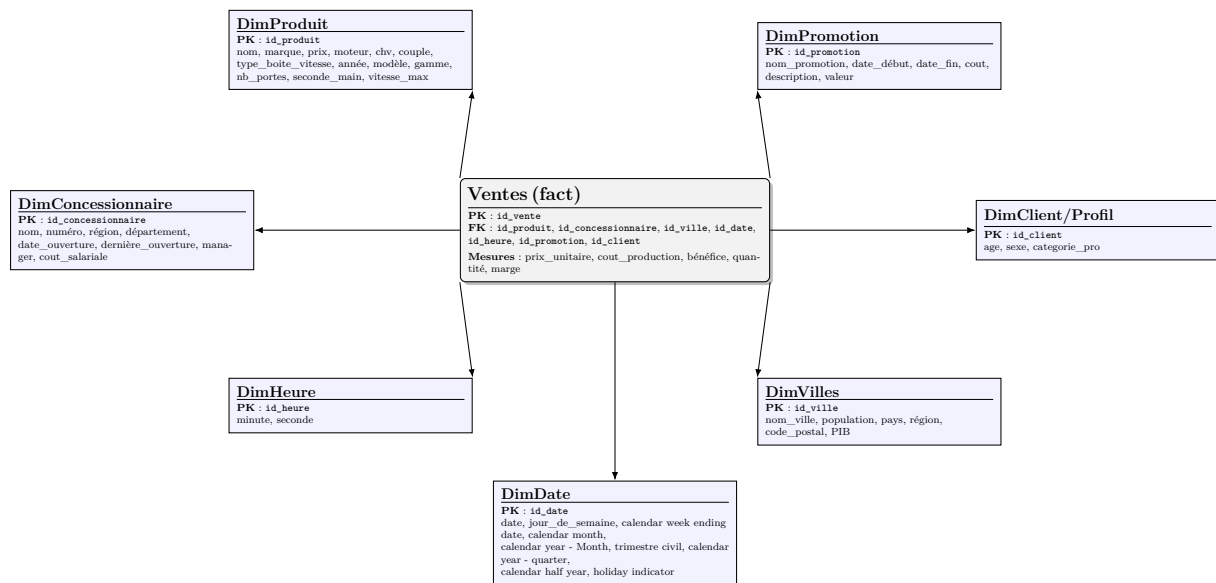


FIGURE 1 – Schéma en étoile des ventes.

6.1 Question 9 - Liste des mesures

- Coût production -> Une mesure additive
- Prix unitaire -> Une mesure non-additive
- Marge -> Une mesure non-additive
- Bénéfice -> Une mesure additive
- Quantité -> Une mesure additive

6.2 Question 10 - Raisonnement pour répondre aux besoins métier

Il est possible de répondre au principal besoin métier. Ça nous permet d'identifier les voitures les plus rentables et d'identifier les modèles clef à mettre en avant et promouvoir.

6.3 Question 11- Pertinence des modèles

TABLE 1 – DimPromotion

id_promotion	nom_promotion	date_début	date_fin	cout (€)	description	valeur
1	Nouvel An	2024-01-01	2024-01-31	200 000	Remises spéciales sur modèles Golf	10 %
2	Toussaint	2024-10-01	2024-11-01	160 000	Offre sur véhicules d'occasion	2 000 €
3	Été VW	2024-07-01	2024-07-31	250 000	Réduction sur gamme électrique	8 %

TABLE 2 – DimProduit

id_produit	nom	marque	prix	moteur	chv	couple	type_boite_vitesse	année	modèle	gamme	nb_portes	seconde_main	vitesse_max
1	Golf 8 GTI	Volkswagen	38000	Essence 2.0	245	370	Automatique	2023	Golf	Sport	5	Non	250
2	A3 35 TFSI	Audi	36500	Essence 1.5	150	250	Manuelle	2022	A3	Compacte	5	Oui	224
3	ID.3 Pro S	Volkswagen	42000	Électrique	—	—	Automatique	2024	ID.3	Électrique	5	Non	160

TABLE 3 – DimConcessionnaire

id_concessionnaire	nom_concessionnaire	numero_concessionnaire	region_concessionnaire	departement_concessionnaire	date_ouverture	derniere_ouverture	manager	cout_salariale
1	VW Toulouse Centre	31001	Occitanie	Haute-Garonne	2012-05-10	2024-03-15	Dupont Marc	1 250 000 €
2	Audi Lyon Sud	69002	Rhône-Alpes	Rhône	2015-09-03	2023-11-10	Martin Julie	980 000 €
3	VW Madrid Norte	28005	Madrid	Madrid	2010-02-21	2024-01-08	Garcia Luis	1 450 000 €

TABLE 4 – DimDate

id_date	date	jour_de_semaine	calendar_week_ending_date	calendar_month	calendar_year_month	trimestre_civil	calendar_year_quarter	calendar_half_year	holiday_indicator
1	2024-01-05	Vendredi	2024-01-07	Janvier	2024-01	T1	2024-Q1	2024-H1	Non
2	2024-07-14	Dimanche	2024-07-14	Juillet	2024-07	T3	2024-Q3	2024-H2	Oui
3	2024-12-31	Mardi	2025-01-05	Décembre	2024-12	T4	2024-Q4	2024-H2	Non

TABLE 5 – DimClient

id_client	age	sexe	categorie_pro
1	34	M	Cadre supérieur
2	27	F	Technicienne
3	52	M	Artisan / commerçant

TABLE 6 – DimVille

id_ville	nom_ville	population	pays	region	code_postal	PIB (€)
1	Toulouse	498003	France	Occitanie	31000	38 200 M€
2	Lyon	522969	France	Rhône-Alpes	69000	74 500 M€
3	Madrid	3266000	Espagne	Madrid	28001	230 000 M€

TABLE 7 – DimHeure

id_heure	minute	seconde
1	15	32
2	47	10
3	5	58

TABLE 8 – Vente(fact)

id_vente	id_produit	id_concessionnaire	id_ville	id_date	id_heure	id_promotion	id_client	prix_unitaire (€)	cout_production (€)	bénéfice (€)	quantite	marge (%)
1	1	1	1	1	1	1	1	38 000	29 000	9 000	1	23.7
2	2	2	2	2	2	2	2	36 500	28 000	8 500	2	23.3
3	3	3	3	3	3	3	3	42 000	34 000	8 000	1	19

6.4 Question 12 - Vérification

La taille du datamart va être de quelques millions de lignes. Par exemple, le nombre total de véhicules vendu est de 9 millions d'unités en 2024. On ne peut pas mettre tout dans un fichier excel car ça serait trop de données pour un seul fichier.

7 Question 13 - Snapshot

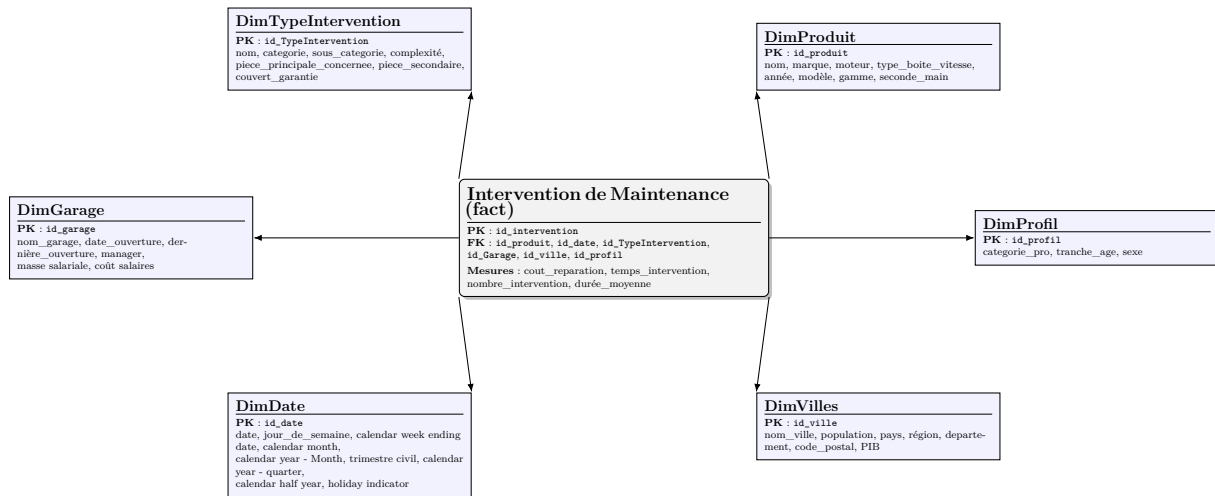


FIGURE 2 – Schéma en étoile des *Interventions de Maintenance*.

7.1 Question 16 - Liste des mesures

- Début intervention -> Une mesure non-additive
- Fin intervention -> Une mesure non-additive
- Temps intervention -> Une mesure additive
- Coût réparation-> Une mesure additive
- Note satisfaction -> Une mesure non-additive

7.2 Question 17 - Raisonnement pour répondre aux besoins métier

Il est possible de répondre au principal besoin métier qui est l'augmentation des recettes car notre updated record met en lumière des coûts additionnels pour chaque véhicule et chaque pièce. Ce qui pourrait potentiellement orienter le groupe Volkswagen vers des alternatives au niveau des pièces pour leur voiture. Quitte à augmenter le coût du véhicule si on peut réduire les frais de réparation.

7.3 Question 18 - Pertinence des modèles

TABLE 9 – Intervention de Maintenance(fact)

id_intervention	id_produit	id_date	id_heure	id_client	id_piece	id_garage	id_statut	cout_reparation	debut_intervention	fin_intervention	temps_intervention	note_satisfaction
1	12	20240115	930	101	5	3	2	350.00	09:00	11:30	150	8
2	8	20240116	1045	102	6	2	3	1200.00	10:00	17:00	420	9
3	14	20240201	1500	103	4	1	1	220.00	14:30	15:30	60	7

TABLE 10 – DimGarage

id_garage	nom_garage	region	departement	date_ouverture	derniere_ouverture	manager	masse_salariale	cout_salaires
1	Garage AutoSud	PACA	Bouches-du-Rhône	2010-05-12	2024-01-10	Jean Dupont	550000	420000
2	Garage NordAuto	Hauts-de-France	Nord	2015-03-01	2023-09-25	Claire Martin	300000	250000
3	Garage ParisCenter	Île-de-France	Paris	2008-07-20	2024-02-15	Marc Leroy	750000	600000

TABLE 11 – DimProduit

id_produit	nom	marque	moteur	type_boite_vitesse	annee	modele	gamme	seconde_main
12	Golf	Volkswagen	Essence	Manuelle	2020	TSI 130	Compacte	Oui
8	A4	Audi	Diesel	Automatique	2022	S-Line	Berline	Non
14	Fabia	Skoda	Hybride	Manuelle	2021	Greenline	Citadine	Oui

TABLE 12 – DimClient

id_client	age	sexe	categorie_pro
101	45	M	Particulier
102	38	F	Professionnel
103	60	M	Particulier

TABLE 13 – DimStatut

id_statut	type
1	En attente
2	En cours
3	Terminée

TABLE 14 – DimPièce

id_piece	nom	type	cout
4	Filtre à huile	Entretien	20
5	Plaquettes de frein	Sécurité	80
6	Courroie de distribution	Moteur	150

TABLE 15 – DimHeure

id_heure	minute	seconde
930	30	00
1045	45	00
1500	00	00

TABLE 16 – DimDate

id_date	date	jour_de_semaine	calendar_month	calendar_year	trimestre_civil	holiday_indicator
20240115	2024-01-15	Lundi	Janvier	2024	T1	0
20240116	2024-01-16	Mardi	Janvier	2024	T1	0
20240201	2024-02-01	Jeudi	Février	2024	T1	0

TABLE 17 – DimVilles

id_ville	nom_ville	population	pays	region	code_postal	PIB
1	Marseille	870000	France	PACA	13000	68000
2	Lille	234000	France	Hauts-de-France	59000	52000
3	Paris	2161000	France	Île-de-France	75000	102000

7.4 Question 19 - Estimation de la taille du datamart

- Source Volkswagen
- Source Garage
- Source Circulation

Il y a environ 18 millions d'interventions réalisées en France. Le groupe Volkswagen englobe en moyenne 15% des véhicules qui sont en circulation, ce qui donne une estimation haute à environ 2 millions d'interventions par an pour le groupe, uniquement en France. On en déduit que la mise en place d'un datamart est justifiée car il y aurait trop de données pour un fichier excel tout seul.

8 Question 14 - Updated record

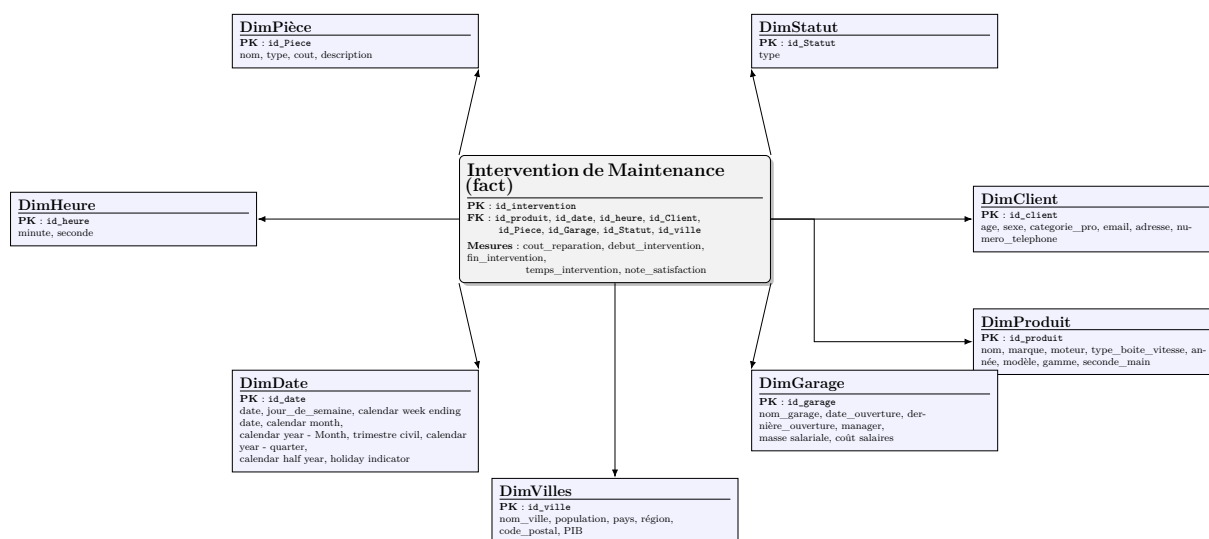


FIGURE 3 – Schéma en étoile des *Interventions de Maintenance* (version avec Statut, Client, Produit, Pièce, Heure, Date, Villes, Garage).

8.1 Question 16 - Liste des mesures

- Coût réparation -> Une mesure additive
- Temps intervention -> Une mesure additive
- Nombre intervention -> Une mesure additive
- Durée moyenne -> Une mesure non-additive

8.2 Question 17 - Raisonnement pour répondre aux besoins métier

Le modèle est un snapshot journalier et nous permet de voir les types d'interventions les plus courants ainsi que les modèles qui bénéficient du plus d'ajustement. Cela permettra au groupe Volkswagen de voir la durée de vie optimale de leur produit et ainsi savoir sur quel catégorie d'intervention, et accessoirement de pièces, ils vont devoir se focaliser.

8.3 Question 18 - Pertinence des modèles

TABLE 18 – DimTypeIntervention

id_TypeIntervention	nom	categorie	sous_categorie	complexité	piece_principale_concernee	piece_secondaire	couvert_garantie
10	Révision complète	Entretien	Périodique	Faible	–	–	Oui
11	Remplacement freins	Sécurité	Freinage	Moyenne	Disques	Plaquettes	Non
12	Diagnostic électronique	Diagnostic	OBD	Élevée	Calculateur	Capteurs	Oui

TABLE 19 – DimProfil

id_profil	categorie_pro	tranche_age	sexe
201	Professionnel	26–35	F
202	Particulier	36–45	M
203	Flotte	46–60	F

TABLE 20 – DimDate

id_date	date	jour_de_semaine	calendar_month	calendar_year	trimestre_civil	holiday_indicator
20240312	2024-03-12	Mardi	Mars	2024	T1	0
20240505	2024-05-05	Dimanche	Mai	2024	T2	0
20240621	2024-06-21	Vendredi	Juin	2024	T2	0

TABLE 21 – DimGarage

id_garage	nom_garage	date_ouverture	derniere_ouverture	manager	masse_salariale / cout_salaires
10	Garage LyonEst	2012-09-10	2023-11-30	Alice Bernard	420000 / 330000
11	Garage ToulouseSud	2018-02-14	2024-03-01	Hugo Perez	310000 / 240000
12	Garage NiceOuest	2009-06-22	2024-01-20	Sarah Cohen	510000 / 410000

TABLE 22 – DimVilles

id_ville	nom_ville	population	pays	region	departement	code_postal / PIB
31	Lyon	522000	France	Auvergne-Rhône-Alpes	Rhône	69000 / 78000
32	Toulouse	498000	France	Occitanie	Haute-Garonne	31000 / 62000
33	Nice	342000	France	PACA	Alpes-Maritimes	06000 / 55000

TABLE 23 – DimProduit

id_produit	nom	marque	moteur	type_boite_vitesse	annee	modele	gamme	seconde_main
21	Tiguan	Volkswagen	Essence	Automatique	2023	R-Line	SUV	Non
22	Q3	Audi	Hybride	Automatique	2021	Quattro	SUV	Oui
23	Octavia	Skoda	Diesel	Manuelle	2019	TDI 150	Berline	Oui

TABLE 24 – Intervention de Maintenance (fact)

id_intervention	id_produit	id_date	id_TypeIntervention	id_Garage	id_ville	id_profil	cout_reparation	temps_intervention	nombre_intervention	durée_moyenne
1001	21	20240312	10	10	31	201	480.00	120	1	120
1002	22	20240505	11	11	32	202	920.00	240	2	120
1003	23	20240621	12	12	33	203	150.00	60	1	60

8.4 Question 19 - Estimation de la taille du datamart

On ne pourrait pas tout mettre dans un fichier excel, on aurait 21 millions de lignes tous les jours, soit environ 8 milliards de lignes par an.

$$90\,000 \times 40 \times 6$$

où :

- 90 000 représente le **nombre de garages Volkswagen officiels** en activité,
- 40 correspond au **nombre moyen de modèles de véhicules différents** observés par jour,
- 6 indique le **nombre de types d'interventions** réalisées par modèle.